



ARTICLE



<https://doi.org/10.1057/s41599-023-02299-1>

OPEN

# Single-item measures of happiness and life satisfaction: the issue of cross-country invariance of popular general well-being measures

Petra Raudenská  <sup>1</sup>✉

Single-item measures of general well-being are increasingly being analysed cross-culturally but without clear evidence of comparability level attainment. The primary objective of this study is to examine the cross-country measurement invariance of the two most common single-item measures—life satisfaction and happiness—across a large number of countries. For this purpose, 45 data sources from large-scale sample surveys conducted between 1976 and 2018 were used. This study presented a novel technique for examining the measurement invariance of individual items and used Bayesian approximation to evaluate the extent of the non-invariance of certain items across nations. The findings revealed that the happiness item's factor loadings and intercepts deviated less, indicating comparability across more countries than the life satisfaction item. It is possible that the construct of happiness is more universally applicable across cultures than that of life satisfaction. However, the item parameters of the survey items varied among several countries in each round of the program, indicating that the observed score means could only be compared between a few participating countries.

<sup>1</sup>Czech Academy of Sciences, Institute of Sociology, Prague, Czech Republic. ✉email: [petra.raudenska@soc.cas.cz](mailto:petra.raudenska@soc.cas.cz)

## Introduction

Survey-based measures of happiness and satisfaction, broadly defined as subjective well-being, are commonly used in empirical psychology, social psychology, and the social sciences. Recently, they have also been widely analysed in economics. Scholars have identified different dimensions of subjective well-being, but three dimensions—cognitive, affective, and eudaimonic—are primarily analysed empirically (Vanhoutte and Nazroo, 2014). The life satisfaction item is assumed to reflect the cognitive component, while the happiness item is believed to reflect the affective component (Strobel et al. 2011). Raudenská (2020) noted that a significant weakness of most subjective well-being studies conducted thus far is their reliance on single-item measures of life satisfaction or happiness as opposed to the more detailed, multi-item measures of well-being (see also Huppert et al. 2009). Although single-item measures lack precision, reliability, and construct validity and do not provide control over measurement errors (Davidov et al. 2018), they remain prevalent in most survey programs.

Furthermore, current comparative studies on subjective well-being often rely on indirect evidence to presume the comparability of data across individuals, nations, cultures, regions, and time periods (Diener, 2009; Bjørnskov, 2010). However, these studies inadequately address the issue of cross-cultural and cross-temporal measurement invariance across extensive samples of countries and time points (Fors and Kulin, 2016; Emerson et al. 2017). Although methodologists have argued that measurement invariance should not be assumed but rather tested empirically, it seems that the responsibility of testing for measurement invariance usually falls on the individual data user (Seddig and Leitgöb, 2018).<sup>1</sup>

Conventional approaches that rely on the latent construct measured by three or more indicators are not viable for single-item measures when performing measurement invariance testing. Revilla and Saris (2011) proposed an alternative approach that utilizes multiple methods in survey design. This method involves asking individuals the same question three times using different methods to achieve the measurement invariance testing of single-item indicators. Measurement invariance testing utilizes a latent construct measured by the same items through different methods. However, many large-scale cross-national social surveys lack a multimethod survey design for practical, financial, and feasibility reasons. Alternatively, they may use a second interviewing method only for unreachable or other respondents.

Therefore, the purpose of this study is to present a simple alternative approach for evaluating the cross-country invariance of the two commonly used single-item measures that gauge overall life contentment and well-being. For this study, almost 2 million participants' data from 45 samples from 1976 to 2018 were analysed. The samples were drawn from the World Values Survey (WVS), International Social Survey Program (ISSP), European Values Study (EVS), European Social Survey (ESS), European Quality of Life Survey (EQLS), and Eurobarometer (EB).

In addition, the current study used the latest Bayesian approximation approach to measure the invariance of well-being research across numerous groups. This approach was selected due to its ability to measure invariance and limited application in the area of well-being research (Raudenská, 2020). Findings regarding the approximation approach can reveal the degree of specific item non-invariance, which is invaluable for this analytical purpose. This study aims to contribute significantly to the literature by examining the comparability of general well-being measures across different countries with the research literature on the psychological properties of measures of subjective well-being. There is a lack of appropriate multi-item instruments available in

cross-national research. Therefore, the proposed innovative approach for single-item measurement invariance testing should advance the discussion and the awareness of measurement invariance in cross-national surveys.

## Theoretical background

**Subjective well-being.** In the literature, subjective well-being is defined as “subjective, self-reported judgments about one’s life” (Diener et al. 2018). Empirical analyses conducted by Vanhoutte and Nazroo (2014) revealed a multidimensional structure of subjective well-being consisting of three distinct dimensions: cognitive, affective, and eudaimonic. The cognitive dimension pertains to one’s overall assessment of life and judgments of life satisfaction (Veenhoven, 2012). Esnaola et al. (2017) posited that individuals who report high life satisfaction tend to perceive their life circumstances as meeting their personal standards. The affective dimension of well-being is characterized by two distinct emotional poles. One end of the spectrum, referred to as positive affect, is often characterized by pleasurable emotions, such as joy or happiness (Diener, 2009). The other end, the negative affect, is commonly associated with uncomfortable feelings, such as anxiety and depression (Nes et al. 2006). The eudaimonic aspect of well-being emphasizes individuals’ fulfillment of their innate potential (Waterman, 1993).

**Measurement of the subjective well-being.** There is a consensus that the construct of general subjective well-being is a multifaceted construct that encompasses more than just happiness or life satisfaction. However, a significant drawback of several surveys conducted so far is their reliance on single-item measures of life satisfaction (“How satisfied are you with your life as a whole?”) and/or happiness (“How happy would you say you are overall?”) rather than multi-item measures (Huppert et al. 2009). The life satisfaction item is believed to reflect the cognitive component, while the happiness item is occasionally considered to reflect the affective component (Strobel et al. 2011).

Most of the large surveys have predominantly used single-item measures to assess overall well-being. These surveys include the WVS, EVS, EB, EQLS, ESS, ISSP, and the Gallup World Poll (2). In addition, popular comparative studies that rank countries according to their level of happiness are often based only on average measurements, without any evidence of the level of comparability achieved (e.g., OECD Better Life Index<sup>3</sup>, World Happiness Report<sup>4</sup>, Quality of Living Rankings from Mercer<sup>5</sup>, and the World Database of Happiness<sup>6</sup>).

However, Veenhoven (2012) argued that “when the same question about happiness is asked twice in an interview, the answers are not always identical and the difference between the response options is often ambiguous. Although responses rarely change from happy to unhappy, changes from ‘very’ to ‘quite’ are quite common.” Single survey items for both linguistic and cultural reasons do not always translate well across countries, tend to be imprecise, and do not have high reliability because the responses are strongly influenced by contextual factors, such as the preceding item, memory bias, desirability bias, or response bias (Billiet, 2003; Saris and Gallhofer, 2007). When only one measure is available, it is not possible to control for random and non-random measurement errors (Davidov et al. 2018). Therefore, single-item measures do not allow for a deeper examination of cross-cultural measurement invariance.

This has led to the development of several multi-item measures of life satisfaction, of which the two best known are Diener et al.’s (1985) five-item Satisfaction with Life Scale (SWLS) (Pavot et al. 1991), which measures the cognitive dimension of subjective

well-being, and the seven-item Personal Well-being Index (PWI), developed by Cummins et al. (2003), which measures life domain satisfaction. Most research on the cross-cultural comparability of subjective well-being scales has been conducted on populations of a specific age (elderly or students), gender, and ethnic group (e.g., Tucker et al. 2006; Clench-Aas et al. 2011; Ponizovski et al. 2013; Vanhoutte and Nazroo, 2014; Tomáš et al. 2015; Dimitrova et al. 2016; Whisman and Judd, 2016; Emerson et al. 2017; Schnettler et al. 2017; Checa et al. 2019; Jang et al. 2017). Very few studies have examined the measurement of cross-national invariance of these multi-item scales using representative national data sources (e.g., Žemojtel-Piotrowska et al. 2017; Jovanović et al. 2018) because they rarely appear in international questionnaires.

Several studies have examined the accuracy of single-item measures of happiness or life satisfaction. Abdel-Khalek (2006) found that the correlation between the single item measuring happiness and the SWLS scale was “highly significant and positive, indicating good concurrent validity.” He also showed that the single item had good convergent validity because it was highly and positively correlated with optimism, hope, self-esteem, positive affect, extraversion, and self-rating of physical and mental health.

Atroszko et al. (2017) found that the correlation between the single-item life satisfaction measure and the SWLS scale was highly significant and positive and that the correlations of both measures with gender, well-being indicators, and personality were not statistically significantly different. Cheung and Lucas (2014) confirmed that the single-item measure and SWLS were similarly correlated with theoretically relevant variables, such as demographics, subjective health, life satisfaction, and affect, in large adult samples from the United States and Germany.

Jovanović (2016) showed that the two scales worked similarly in three samples of Serbian adolescents. Jovanović and Lazić (2020) confirmed that they were highly correlated across six Serbian samples of different ages. Furthermore, Jovanović and Brdar (2018) showed that the SWLS was strongly correlated with a single-item measure and revealed similar correlations with other constructs in Austria, Bosnia and Herzegovina, Croatia, Montenegro, and Serbia. Fonberg and Smith (2019) found positive correlations with single-item and multi-item measures of life satisfaction, positive personality (self-efficacy, self-esteem, and optimism), positive affect, and happiness, and negative associations with negative affect and anxiety/depression.

The aforementioned authors concluded that single-item measures of well-being performed similarly to the multiple-item scales and that social scientists would receive a similar response to substantive questions regardless of which measure they used. Previous studies have found extremely high response rates for single-item life satisfaction questions (Diener et al. 2013). “This suggests that most people do not have difficulty understanding these questions. The simplicity of single-item life satisfaction questions is also supported by findings that the percentage of participants answering ‘don’t know’ to these questions is less than 1% in most countries” (Veenhoven, 2010) and that it takes only two seconds for most people to answer them (Oishi, 2012).

Unfortunately, the cultural invariance of general well-being measures is rarely empirically assessed in cross-national studies, and typically, researchers simply assume that the reported levels of happiness or life satisfaction are comparable across countries (Jovanović and Brdar, 2018). No previous research has investigated whether general well-being measures have the same meaning for individuals in different countries, mainly because they are single items and there is no analytical tool to measure them accurately. The present study aims to fill this gap in well-being research by testing the cross-country measurement invariance of the two most commonly used single-item measures

of general life satisfaction and happiness across a large set of countries using 45 data sources from large sample surveys from 1976 to 2018.

**Testing the measurement invariance of single items.** Measurement invariance, by definition, is a situation in which the operationalization of a construct results in the measurement of completely identical characteristics under different circumstances in which a given phenomenon is studied (Horn and McArdle, 1992). By different circumstances, we mean different measurement times, measurement of different populations/groups, or the use of different data collection methods. To date, the methods for testing measurement invariance have received much attention in the literature (Leitgöb et al., 2022), but the best-known applied technique is still multiple-group confirmatory factor analysis (MGCFA), which is based on a construct (a latent variable) that can be measured by multiple indicators (observed variables).

A clear advantage of the MGCFA is its ability to test the comparability of attitude scales across different groups and different waves of the survey, and the possibility of more in-depth analysis of different levels of measurement invariance. The procedure for testing particular levels of measurement invariance using the MGCFA has been described in detail by Steenkamp and Baumgartner (1998) and Vandenberg and Lance (2000). The essence of this testing is to verify the similarity of the factor structure of the measurement model across groups. The MGCFA measurement invariance testing is a hierarchical, stepwise process. This strategy is characterized by the increasing restriction placed on the model under test. Increasing constraint refers to an increasing number of model parameters (i.e., factor loadings, intercepts, measurement errors, etc.) that are required to be identical across the groups under test, even though the variables and the relationships between them remain unchanged in the model. Thus, the models defined are tested for consistency with the research data. However, conventional measures based on a latent variable that can only be measured by at least three observed variables cannot be computed for single-item measures.

Revilla and Saris (2011) proposed a possible approach to test the measurement invariance of individual indicators. They recommended the use of multiple methods in a survey design in which the same question is asked several times using (at least three) different methods (e.g., face-to-face, paper, and web surveys). The repetition should take place after at least 20 min to avoid memory effects (Saris and Van Meurs, 1990). The measurement invariance test then uses the latent construct measured by these (same) items, measured by different methods. In this latent construct, different methods are used to measure the same concept rather than different items. However, this proposal for testing the measurement invariance of individual indicators is theoretical because, to the best of our knowledge, it has not yet been used in any published methodological study. This approach requires a multi-method survey design, which is not common in cross-national surveys, thus making it difficult to apply in practice.

In this study, we proposed an innovative (yet basic) approach for assessing the measurement invariance of single items using other subjective well-being measures available in cross-national questionnaires to create the best possible general well-being construct based on (not only) a high correlation among available items—that is, the approach of synthesizing multi-item instruments from several single-item variables. In almost all the selected large-scale sample surveys, we found two single items measuring life satisfaction and happiness with high correlations. We mostly selected the single-item measure of subjective health and/or life domain satisfaction (i.e., satisfaction with work, family, or

democracy), which is the third most highly correlated item with the life satisfaction scales.

An important discussion to be addressed here is whether it is possible to capture general well-being as a latent concept using three different single items rather than a standardized scale with very similar items. From a theoretical point of view, especially in this particular case, we think that this is possible. It is common to capture different facets of well-being in multi-item scales, such as the SWLS, PWI, or ESS well-being module (see Huppert et al. 2009), including an emphasis on personal and social well-being, cognitive and affective dimensions of life satisfaction, mental and vital well-being, and so on. These are well-established instruments that capture multiple dimensions of well-being. In our analysis, the satisfaction item captures the cognitive aspect more, the happiness item captures the affective aspect more, the life domain satisfaction item captures personal and social well-being, and the subjective health item can reflect mental well-being.

However, it should be noted that, as the aim is to test the invariance of individual items across countries, the extracted factor becomes an important benchmark. If this benchmark represents a contaminated measure of the latent construct, the results will be compromised. This could lead to a biased construct covariance in which only three individual items could overlap or, in other words, have random correlations in a synthetic latent variable. As a result, the measurement invariance test could produce misleading results. This potential pitfall is statistically explored in the Results Section.

To assess the extent of item non-invariance, we used the latest Bayesian approximation approach. Unlike the traditional exact measurement invariance approach, which is often used for configural, metric, and scalar invariance tests (e.g., Meredith, 1993; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000; Davidov et al. 2014; Kim et al. 2017), the Bayesian approximation approach replaces exact equality constraints on factor loadings and intercepts with the requirement that all parameters are approximately equal (Muthén and Asparouhov, 2013; Van de Schoot et al. 2013). In the exact measurement invariance approach, the differences between factor loadings (metric invariance) or intercepts (scalar invariance) are constrained to zero across groups. By contrast, in the Bayesian approximation approach, the differences among these parameters are assumed to be close to zero (Muthén and Asparouhov, 2013). However, the differences are kept to a minimum to ensure that the concepts remain approximately comparable. This option prevents the model fit from being subjected to an unreasonable assumption of identical constraints that do not reflect the original intention of the researchers. This approach has been successfully applied in numerous recent comparative studies (e.g., Bujacz et al. 2014; Ciecuch et al. 2014; Davidov et al. 2015; Zercher et al. 2015; Ciecuch et al. 2018; Davidov et al. 2018; Seddig and Leitgöb, 2018; Raudenská, 2020).

The approximation approach using a Bayesian framework typically requires constraining the mean difference between parameters to be zero and the variance of the parameters to be greater than zero but sufficiently small (i.e., the prior variance). The size of the prior variance reflects the level of approximation: the smaller the variance of the difference, the more restrictive the model, and the more similar it is to an exact measurement invariance model (Ciecuch et al. 2018). Conversely, a (very) large variance for the parameter difference reflects non-invariance for that parameter.

Van de Schoot et al.'s (2013) simulation studies suggest that a variance of 0.05 can be allowed without risking invalid latent mean comparison inferences. However, Pokropek et al. (2020) argued that higher prior variances, such as 0.05 or 0.1, go beyond the definition of “small between-group discrepancies” and imply

high levels of differences between item parameters. The authors recommended starting with the simplest model (with a prior equal to zero) and then gradually increasing the prior (e.g., 0.001, 0.005, 0.01, 0.025, 0.05) until a significant improvement in model fit is achieved. They also suggested what threshold of model fit indices should be used to decide where to stop—that is, to determine that the correct prior has been reached and does not need to increase further.

The deviance information criterion (DIC) and the Bayesian information criterion (BIC) are the most popular criteria for Bayesian model selection and model comparison, and they may be preferred in Bayesian structural modeling. The DIC measures the posterior predictive error by penalizing the fit of a model according to its complexity, which is determined by the effective number of parameters (Seddig and Leitgöb, 2018). The model with the lowest DIC (and BIC) value is preferred.

Two measures of fit are typically used to determine whether approximate invariance is present (Muthén and Asparouhov, 2013; Kim et al. 2017): the posterior predictive probability value (PPP) and the 95% credibility interval (CI). The Bayesian model fits the data well when the PPP is not significant and the CI contains zero. In addition, a PPP value of 0.5 or greater indicates that the model fits the data very well (Muthén and Asparouhov, 2013; Van de Schoot et al. 2013). Pokropek et al. (2020) suggested that when searching for the correct prior variance, a BIC improvement (i.e., decrease) of 20 or more justifies a higher prior choice, whereas a smaller change in BIC suggests not increasing the prior variance when evaluating approximate measurement invariance. In their study, the recommended DIC threshold was 14 or higher, and the suggested PPP threshold was 0.025 for medium and large surveys of 24–30 countries, with more than 1500 participants in a country.

The Mplus software package provides researchers with another type of output to assess specific item non-invariance—the difference output. The difference output shows the mean loading/intercept across groups and the amount by which each group-specific loading/intercept deviates from this value (Lek et al. 2018). In other words, this part of the output lists all parameters that are too different in each group. Based on this list, researchers can conclude which countries and/or items are approximately invariant and which are not (Muthén and Asparouhov, 2013; Ciecuch et al. 2018). The possibility of using a mean of single items measuring happiness/satisfaction for comparison across countries, as many comparative studies do, requires full scalar invariance, which means that respondents with the same score on the construct have the same expected response. For this reason, we attempted to identify the deviating loadings/intercepts of individual items measuring life satisfaction and happiness in the difference output and assess the countries in which the items were approximately invariant and the countries in which they were not.

## Data and methods

**Data.** The analysis was based on 45 data samples from large-scale, nationally representative, cross-national surveys: seven rounds of the WVS 1981–2017; four rounds of the ISSP 2002, 2011, 2012, and 2017; five rounds of the EVS 1981–2017; nine rounds of the ESS 2002–2018; four rounds of the EQLS 2003–2016; and 16 rounds of the EB 1976–1979, 1982–1986, 1993–1995, and 1998–2001. They were organized from 1976–2018 and included data for the full set of more than 2,000,000 individuals from around the world (see Tables S2–S7 in the Supplementary Information for more details). The data are freely accessible from the ESS website ([www.europeansocialsurvey.org](http://www.europeansocialsurvey.org)), the WVS website (<http://www.worldvaluessurvey.org/wvs.jsp>), and the

GESIS ZACAT archive (<https://zacat.gesis.org/webview/>) after user registration. Further information on the data collection, the sampling procedure, the questionnaires, and other methodological documentation is also available on these websites.

As some research programs use different modes of data collection over time, a number of respondents had to be excluded. Depending on the presence of an interviewer, respondents may interpret the same question and/or response category differently and may give a different answer simply because of the way the question was presented (de Beuckelaer and Lievens, 2009; Hox et al. 2015). Previous studies have generally found that scalar equivalence is more common between different self-administered modes (e.g., pencil vs. online) and between different interviewer-administered modes (e.g., face-to-face vs. telephone) than between self-administered and interviewer-administered modes (e.g., online vs. face-to-face) (Cernat and Revilla, 2021; Sakshaug et al. 2022).

Therefore, in the datasets, we retained the respondents who underwent an interviewer-assisted interview (paper-and-pencil, face-to-face, telephone, or computer-assisted), and self-completed questionnaires were filtered out. In the WVS 2017, we excluded interview modes other than assisted interviewing, such as self-administered mail survey or online survey.<sup>7</sup> In the ISSP, we also excluded other interview modes, such as self-completion (arriving with the interviewer, mailed to the respondent, CASI, or web questionnaire).<sup>8</sup> In the EVS 2017, we excluded a self-administered web survey.<sup>9</sup> Some countries were excluded from the analysis because they contained cases with missing data for all variables. Other missing data were treated using Bayesian modelling, which treats missing data similar to the full information maximum likelihood (Asparouhov and Muthén, 2010).

**Instrument.** The underlying data came from the main questionnaires, which contained several items related to subjective well-being. The first question of interest was the single-item happiness measure “Taking all things together, would you say you are [how happy would you say you are]?” on self-rating scales with different options, mainly the verbal Likert scale ranging from (1) “not at all happy” to (4) “very happy.” In all rounds of the ESS and EQLS surveys, a 10-point numerical scale was used. In the ISSP 2002, 2011, and 2012, a different wording was chosen: “If you were to consider your life in general [these days], how happy or unhappy [would] you say you are, on the whole?” The response options ranged from (1) “completely unhappy” to (7) “completely happy” (for the exact wording of specific questions in specific data sources, see Table S1 in the Supplementary Information). In the ISSP 2011 and 2017, another item was used to measure happiness: “During the past four weeks, how often have you felt unhappy and depressed?”, with response options ranging from (1) “never” to (5) “very often.” This single-item happiness measure was missing from the EB 1993–2001.

The second question of interest was the single-item life satisfaction measure “All things considered, how satisfied are you with your life as a whole [would you say you are with your life] these days [nowadays]?” The respondents were required to use self-rating scales with different options, mainly a 10-point numerical scale ranging from (1) “dissatisfied” to (10) “satisfied,” except for the ISSP 2017, which has a seven-point verbal scale. This single-item measure of life satisfaction was missing in the ISSP 2002, 2011, and 2012.

As previously mentioned, to test for measurement invariance, we needed to construct the general well-being measured by at least three items. Based on the high correlation, which is both theoretically and empirically supported, we chose either a single-item general health measure (Diener and Chan, 2011) or one of

the life domain satisfaction items (i.e., satisfaction with job, family, or democracy). The other questions of interest were “All in all, how would you describe your state of health these days?” and “In general, would you say your health is [how is your health]?” The response options ranged from (1) “very poor/poor/very bad” to (5) “very good/excellent.” For the life domain satisfaction items, the respondents were asked, “All things considered, how satisfied are you with your job/family/democracy work in your country?” on self-rating scales with different options, mainly a seven-point verbal scale or a 10-/11-point numerical scale (see Table 1 for the specific configuration of a general well-being construct in specific data samples).

The correlations between items measuring the corresponding life satisfaction, happiness, life domain satisfaction, and subjective general health were quite high, ranging from 0.3 to 0.6 across all datasets.

**Analytical strategy.** All analyses were performed using the software package Mplus 7.4 (Muthén and Muthén, 1998–2017). First, the descriptive statistics (means, standard deviations, minimum, and maximum) were calculated, the data were checked for normality (Shapiro–Wilk normality test, skewness, and kurtosis), some scales were reverse-coded, and all items were standardized according to Pokropek et al.’s (2020) recommendation when dealing with the different distributions of the tested items. We created a z-score, or a standard score, to standardize scores on the same scale by dividing the deviation of a score by the standard deviation in a dataset—in this case, in each country and round separately.

Second, the Bayesian approximation approach was used to test the invariance of the general well-being construct. This procedure is applicable to latent variable models with categorical non-normal data (Muthén and Asparouhov, 2013) and is suitable for a large number of groups. The prior mean of the differences between loadings and intercepts across countries was set to 0, and the prior variance was set incrementally to 0.000, 0.001, 0.005, 0.01, 0.025, and 0.05. The evaluation of the model was based on the PPP, CI, and suggested thresholds for the correct prior variance (i.e., a BIC improvement of 20 or greater and a DIC improvement of 14 or greater; the threshold for PPP was 0.025).

We chose the strategy of counting the number of iterations in the Mplus inputs when using BITERATIONS to determine the minimum number of total iterations and then find the moment when the potential scale reduction (PSR) meets the convergence criterion. As the number of iterations increases, the computational time, memory, and performance requirements of the hardware also increase. For example, in this particular case (i.e., 45 datasets), there are 300 steps of the Bayesian analysis. When the number of iterations reaches 5000, an analysis step takes about 5 min, when the number of iterations reaches 20,000 or 50,000, the time increases from 15 min to 2 h, depending on the size of the data file. For this reason, when it was not possible to use BITERATIONS, we used a fixed number of iterations with the lowest PSR value of around 1.0. Finally, we evaluated the Mplus difference output results to assess the non-invariance of specific items in different countries. See Table S14 in the Supplementary Information for examples of the Mplus scripts for the Bayesian approximate measurement invariance of the general well-being latent construct in different surveys.

## Results

First, the single-sample CFA analysis for the general well-being construct (using the happiness, satisfaction, and subjective health/life domain satisfaction items) was performed separately in each country and round, showing an acceptable fit (not shown here).



**Table 2 Mplus difference output—countries with significant deviations of item loadings and intercepts relative to the average across all countries; World Values Survey 1981–2017.**

Survey/item	stflife (10-point numerical scale)	happy1 (4-point verbal scale)	health1 (5-point verbal scale)
	LAM & NU	LAM & NU	LAM & NU
WVS 1981	<b>South Korea</b>		Japan
WVS 1990		Chile	<b>South Korea</b>
WVS 1995	Lithuania, New Zealand, Colombia, Mexico		<b>Germany</b> , New Zealand, Ukraine, Argentina
WVS 2000	Canada, Spain, Indonesia, Singapore, Iraq, Israel, Morocco, Saudi Arabia, <b>Egypt</b> , Nigeria, Zimbabwe, Uganda		Canada, Montenegro, Spain, Vietnam, Iraq, Morocco, <b>Egypt</b> , Nigeria, Zimbabwe
WVS 2005	Ghana, Egypt, <b>South Korea</b>	Russia	<b>Cyprus</b> , Sweden, Hungary, Russia, Turkey, Iran, Jordan, Ghana, Mexico, Malaysia
WVS 2012	<b>Germany</b> , Netherlands, Slovenia, United States, Sweden, Kyrgyzstan, Algeria, Kuwait, Nigeria, Pakistan, Rwanda, Japan, Malaysia, Thailand	Algeria, Armenia	<b>Germany</b> , Cyprus, Kyrgyzstan, Kuwait, Nigeria, Libya, Rwanda, Japan, Trinidad and Tobago
WVS 2017	<b>Germany, Cyprus</b> , Greece, Serbia, United States, Ethiopia, Tunisia, Bangladesh, China, Thailand, Iran, Iraq	<b>Cyprus</b> , Serbia, Malaysia, Iran	Greece, Romania, Lebanon, Tunisia, Turkey, <b>Egypt</b> , China, <b>South Korea</b>

*Data source: World Values Survey 1981–2017; LAM (factor loadings), NU (intercepts/thresholds); countries in bold are those with more frequent deviated item parameters, the empty box indicates the research year where all countries were invariant.*

All factor loadings were significantly high, exceeding 0.4 for all items.

As we should examine whether the extracted factor, based on three single-item measures, is a contaminated measure of the latent construct, which would lead to biased construct covariance (see Section “Testing the measurement invariance of single items”), we analysed the correlations between the error terms of the satisfaction and happiness items. If our extracted factor is indeed a general well-being factor, then the error terms of these two items should be uncorrelated. However, if the third single item (e.g., subjective health, and life satisfaction domains) is biasing the latent construct of general well-being, then some of the covariance between the satisfaction and happiness items have been pushed into the residuals and become visible as correlated error terms.

This was tested with a (a) configural invariance model WITH correlations between the error terms of the satisfaction and happiness items and a (b) configural invariance model WITHOUT these correlations using multiple-group CFA. The results showed that the configural model WITH correlations did not terminate normally and could not be computed in any survey round, while the configural model WITHOUT these correlations fit the data much better. Therefore, we can assume that the latent construct of general well-being is not contaminated by random intersections.

Second, the proposed model was simultaneously tested across countries in a specific dataset using Bayesian modelling to analyse the approximate measurement invariance. Tables S8–S13 in the Supplementary Information present the global fit statistics for the approximate scalar invariance test of the general well-being construct across the countries in each round of specific cross-national surveys, separately, with different prior variances. The models with low prior variances of 0.000 and 0.001 showed a good fit with the data, and the prior variances remained acceptably low. Choosing a more liberal prior did not sufficiently improve the fit of the model. Thus, the approximate scalar invariance of the general well-being construct, which is considered necessary for comparing latent means across groups, was established in all cases.

However, the study aimed to assess the extent of specific single-item non-invariance. The Mplus difference output showed that most of the item loadings and/or intercepts were usually approximately invariant across most of the countries in each round (Tables 2–7). This means that, in most cases, it is possible

to use the items for analyses comparing factor/item covariances, unstandardized regression coefficients across samples, and item means.

However, the results also showed that some item parameters of both single-item measures of happiness and life satisfaction were scalar non-invariant in several countries and survey programs. Comparing the approximate invariance results, the single-item happiness measure produced better results than the single-item life satisfaction measure in 13 cases (out of 45 data samples). By contrast, the life satisfaction measure outperformed the happiness measure in only three cases, and the results were equal in 19 cases.

For example, in the first round of the WVS 1981, the happiness measure with a four-point verbal scale was scalar invariant in all countries, while the life satisfaction measure with a 10-point numerical scale was scalar non-invariant in one country (i.e., South Korea). In the most recent round of the WVS 2017, the loadings/intercepts of the happiness item deviated in four countries. Conversely, the loadings/intercepts of the life satisfaction measure deviated in 12 countries. In general, the happiness measure with a four-point verbal scale was approximately scalar invariant in 9–58 countries in the WVS 1981–2017, while the life satisfaction measure with a 10-point numerical scale was approximately invariant in only 8–45 countries (Table 2). The degree of non-invariance found for the subjective health measure with a five-point verbal scale was similar to that of the life satisfaction measure; it was approximately invariant in only 8–51 countries.

In the ISSP 2002–2017, it was not possible to directly compare the results of the measurement invariance for the two single-item measures because they were not asked at the same time in the main questionnaire. However, the happiness measure with a seven-point verbal scale had a slightly higher number of invariant parameters than the subjective health and life satisfaction items (Table 3).

In the EVS 1981–2017, the single-item happiness measure (with a four-point verbal scale) showed slightly better results of approximate measurement invariance than the life satisfaction measure (with a 10-point numerical scale); it was approximately invariant in 16–47 countries (Table 4). The life satisfaction measure and the subjective health item with a five-point verbal scale were only approximately invariant in 16–38 and 15–31 countries, respectively. The subjective health item showed much less comparability across countries in these rounds.

**Table 3 Mplus difference output—countries with significant deviations of item loadings and intercepts relative to the average across all countries; International Social Survey Programme 2002, 2011, 2012, 2017.**

Survey/item	happy2 (7-point verbal scale) LAM & NU	happy_feel (5-point verbal scale) LAM & NU	stflife (7-point verbal scale) LAM & NU	stfjob (7-point verbal scale) LAM & NU	stffam (7-point verbal scale) LAM & NU	health2 (5-point verbal scale) LAM & NU
ISSP 2002		-	-	Hungary		-
ISSP 2011			-	-	-	<b>United States</b>
ISSP 2012		-	-	-	Bulgaria, <b>China</b>	Mexico
ISSP 2017	-	<b>China</b> , India, Russia	Taiwan, Croatia	-	-	India, Taiwan, <b>United States</b>

Data source: International Social Survey Programme 2002, 2011, 2012, 2017; LAM (factor loadings), NU (intercepts/thresholds); countries in bold are those with more frequent deviated item parameters, the empty box indicates the research year where all countries were invariant.

**Table 4 Mplus difference output—countries with significant deviations of item loadings and intercepts relative to the average across all countries; European Values Study 1981–2017.**

Survey/item	happy1 (4-point verbal scale) LAM & NU	stflife (10-point numeral scale) LAM & NU	stfjob (10-point numeral scale) LAM & NU	health1 (5-point verbal scale) LAM & NU
EVS 1981			-	<b>Canada</b>
EVS 1990	<b>Canada</b>	<b>Canada</b>	-	
EVS 1999	<b>Italy</b> , Turkey		<b>Italy</b> , Russia	-
EVS 2008		<b>Germany</b> , Greece, Netherlands, <b>Norway</b> , Romania, Turkey, Macedonia, Kosovo, Azerbaijan	-	<b>Germany</b> , Cyprus, Netherlands, <b>Norway</b> , Slovakia, Turkey, Macedonia, Kosovo, Azerbaijan, <b>Great Britain</b> , France, <b>Iceland</b> , Finland, <b>Italy</b> , Lithuania, Bosnia Herzegovina
EVS 2017		Austria, <b>Norway</b> Slovakia, Romania, <b>Great Britain</b> , <b>Iceland</b>	-	<b>Germany</b> , <b>Norway</b> , Slovakia, Bulgaria, Armenia, <b>Great Britain</b> , <b>Iceland</b> , Lithuania

Data source: European Values Study 1981–2017; LAM (factor loadings), NU (intercepts/thresholds); countries in bold are those with more frequent deviated item parameters, the empty box indicates the research year where all countries were invariant.

Similarly, in the ESS 2002–2018, the happiness measure (with a 10-point numerical scale) showed slightly better results (from 22 to 27 invariant countries) than the life satisfaction measure (with a 10-point numerical scale) (Table 5). The subjective health item with a five-point verbal scale was approximately invariant in only 16–17 countries.

The differences between these two indicators with a 10-point numerical scale were negligible in the EQLS, in which the happiness and life satisfaction measures showed similar comparability across countries (from 28 to 33 invariant countries) from 2003 to 2016 (Table 6). The subjective health item with a five-point verbal scale was only approximately invariant in 24–27 countries.

In the EB 1976–1986, the measures of happiness (with a four-point verbal scale) and life satisfaction (with a 10-point numerical scale) showed similar results, being invariant in 9–12 countries (Table 7). In the EB 1993–2001, the results of the invariance of the life satisfaction measure were even worse than those of the items measuring satisfaction with democracy (with an 11-point numerical scale), which were approximately invariant in 13–16 countries.

In addition, the pattern of violations of item measurement invariance across countries was not random as certain countries tended to have violations of measurement invariance for multiple surveys and instruments. For example, in the WVS, some countries showed systematic deviations in the item loadings and/or intercepts of the happiness or life satisfaction measures across rounds (e.g., South Korea, Germany, Egypt, and Cyprus). In the ISSP, India and the United States showed repeated deviations in the happiness or life satisfaction item parameters across rounds. In the EVS, Canada, Italy, Germany, Norway, the United Kingdom, and Iceland were excluded from the invariant countries and showed some systematic patterns of deviations. In the ESS,

Hungary, Portugal, Denmark, the United Kingdom, and Estonia were excluded, but these countries showed frequent deviations in the item loadings/intercepts of the subjective health single item. The same pattern held for the EQLS, although Bulgaria showed frequent deviations in the item loadings/intercepts of the subjective health single item. In the EB, France, Belgium, the Netherlands, and Germany systematically deviated.

To consider the implications of the violations of measurement invariance of item parameters for researchers and their analyses, it is important to examine the extent to which the latent and observed means of individual items produce different country rankings to capture potential biases in analyses with cross-national data—that is, to show how much the true mean differences between countries are over- or underestimated. Figures 1–5 show the differences between the observed means for the individual happiness items and the latent means based on the approximation approach for the general well-being construct for all countries in the latest round of specific cross-national surveys. The black colour highlights the observed scores that should not be used for comparison due to the item non-invariance calculated by the Mplus difference output discussed above. For the other observed scores, the approximate scalar invariance held, and they could be compared.

In the WVS 2017, ISSP 2017, and EVS 2017, somewhat stronger differences between country rankings were found, with correlations between the latent and observed score mean rankings ranging from 0.1 to 0.3 (Figs. 1–3). Several observed scores should not be used for comparison due to item non-invariance. This means that using the means based on the observed scores could lead to erroneous conclusions about country rankings on happiness. The main reason why the results of the comparison of country rankings are much worse in the ISSP 2017 may be the



**Table 5 Mplus difference output—countries with significant deviations of item loadings and intercepts relative to the average across all countries; European Social Survey 2002–2018.**

Survey/item	happy1 (10-point numeral scale) LAM & NU	stflife (10-point numeral scale) LAM & NU	health2 (5-point verbal scale) LAM & NU
ESS 2002		<b>Hungary, Portugal</b> Israel	<b>Hungary, Portugal, United Kingdom</b> , Luxembourg, Finland
ESS 2004		<b>Portugal</b>	<b>Denmark, United Kingdom</b> , Spain, Turkey,
ESS 2006			<b>Denmark</b> , Ireland
ESS 2008		Cyprus	Ireland, Turkey
ESS 2010			<b>Hungary, Portugal</b> , Bulgaria, Ukraine <b>Denmark, United Kingdom</b> , Norway, Greece
ESS 2012			<b>Hungary, Portugal</b> , Bulgaria, Ukraine Belgium, Spain, Ireland, <b>Estonia</b> , Lithuania, Cyprus, Israel
ESS 2014		<b>Hungary</b>	<b>Estonia</b>
ESS 2016			<b>Hungary</b> Belgium, Spain, Ireland, <b>Estonia</b> , Lithuania Israel
ESS 2018			Cyprus

Data source: European Social Survey 2002–2018; LAM (factor loadings), NU (intercepts/thresholds); countries in bold are those with more frequent deviated item parameters, the empty box indicates the research year where all countries were invariant.

**Table 6 Mplus difference output—countries with significant deviations of item loadings and intercepts relative to the average across all countries; European Quality of Life Survey 2003–2016.**

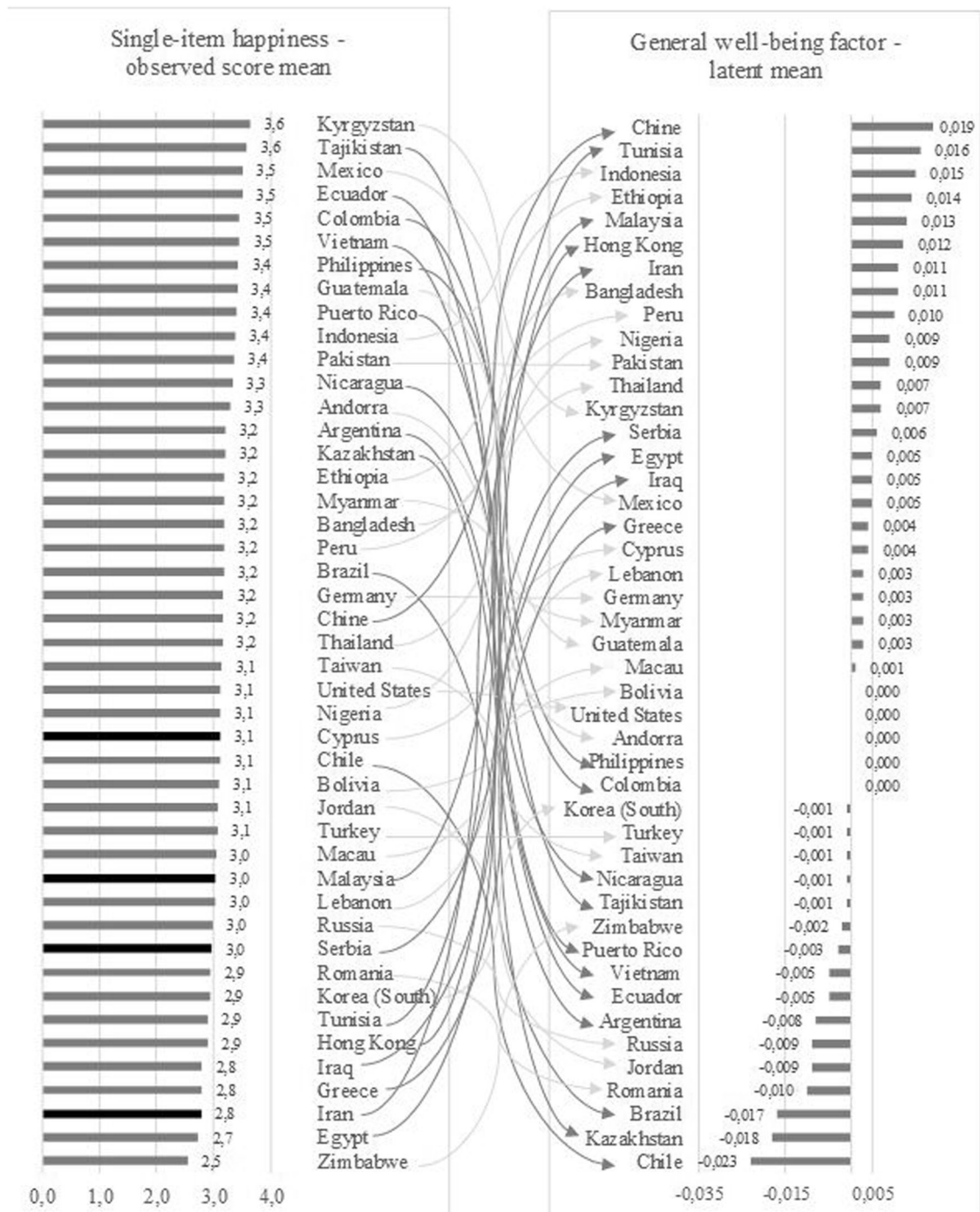
Survey/item	happy1 (10-point numeral scale) LAM & NU	stflife (10-point numeral scale) LAM & NU	health2 (5-point verbal scale) LAM & NU
EQLS 2003			France, Luxembourg, <b>Bulgaria</b> , Greece
EQLS 2007			
EQLS 2011		Turkey	Austria, <b>Bulgaria</b> , Spain, France, Estonia, Macedonia, Romania,
EQLS 2016			Ireland, Italy, <b>Bulgaria</b> , Estonia, Macedonia, Turkey, Montenegro

Data source: European Quality of Life Survey 2003–2016; LAM (factor loadings), NU (intercepts/thresholds); countries in bold are those with more frequent deviated item parameters, the empty box indicates the research year where all countries were invariant.

**Table 7 Mplus difference output—countries with significant deviations of item loadings and intercepts relative to the average across all countries; Eurobarometer 1976–1979, 1982–1986, 1993–1995, 1998–2001.**

Survey/item	happy1 (4-point verbal scale) LAM & NU	stflife (10-point numeral scale) LAM & NU	stfdem (11-point numeral scale) LAM & NU	stfdeu (11-point numeral scale) LAM & NU
EB 1976				-
EB 1977			<b>France, Belgium, Netherlands</b>	-
EB 1978			<b>Netherlands</b>	-
EB 1979			<b>France, Netherlands</b>	-
EB 1982			<b>Netherlands</b>	-
EB 1983			<b>Netherlands, Germany</b> , Italy	-
EB 1984			<b>Netherlands</b> , Denmark	-
EB 1985			<b>Netherlands</b> , Portugal	-
EB 1986				-
EB 1993	-			Norway
EB 1994	-	<b>France, Belgium, Germany</b> , Luxembourg, Norway		United Kingdom
EB 1995	-	<b>Germany</b>		Norway
EB 1998	-	<b>Belgium, Germany</b>		
EB 1999	-	<b>Germany</b> , Ireland, Greece, Spain		
EB 2000	-	<b>Germany</b> , Greece, Portugal		
EB 2001	-			<b>France</b> , Portugal

Data source: Eurobarometer 1976–1979, 1982–1986, 1993–1995, 1998–2001; LAM (factor loadings), NU (intercepts/thresholds); countries in bold are those with more frequent deviated item parameters, the empty box indicates the research year where all countries were invariant.



**Fig. 1 Latent and observed score means differences for general well-being factor and single-item measures of happiness/World Values Survey 2017.**  
 Data source: World Value Survey 2017; author's figure. Note: Single-item question: Taking all things together, would you say you are [how happy would you say you are]? (1 not at all happy-4 very happy). Latent means were estimated in approximate invariant model (with a prior variance of 0.001). Pearson correlation coefficient between rankings of countries result from latent and observed score means comparison. single-item happiness score = -0.1, single-item life satisfaction score = -0.2. Observed scores that should not be used for comparison due to non-invariance are highlighted in black. Single arrows indicate the different ranking of a particular country by the observed score mean of happiness versus the ranking of country by latent mean. Light grey arrows indicate a small change in the ranking of a given country, dark grey arrows indicate a significant change in the ranking of a given country.

different wording of the happiness item (“In the last four weeks, how often have you felt unhappy and depressed?”).

The differences in country rankings were smallest in the ESS 2018, with the correlation between the latent and observed means being 0.6 (Fig. 4). In the figure, the individual arrows in light grey indicate a small change in the ranking of a given country by the country by the observed score mean of happiness compared with

the country’s ranking by the latent mean (i.e., latent country mean has shifted by one to four places.). This means that Switzerland ranked first based on the observed mean score and was one place lower based on the latent mean. The country’s true mean might have been slightly overestimated.

In the EQLS 2016, we also found strong differences in the country rankings, with a correlation between the latent and



**Fig. 2 Latent and observed score means differences for general well-being factor and single-item measures of happiness/International Social Survey Programme 2017.** Data source: *International Social Survey Programme 2017*; author's figure. Note: Single-item question: During the past 4 weeks how often have you felt unhappy and depressed? (1 very often –5 never). Latent means were estimated in approximate invariant model (with a prior variance of 0.001). Pearson correlation coefficient between rankings of countries result from latent and observed score means comparison. single-item happiness score =  $-0.04$ , single-item life satisfaction score =  $-0.2$ . Observed scores that should not be used for comparison due to non-invariance are highlighted in black. Single arrows indicate the different ranking of a particular country by the observed score mean of happiness versus the ranking of country by latent mean. Light grey arrows indicate a small change in the ranking of a given country, dark grey arrows indicate a significant change in the ranking of a given country.

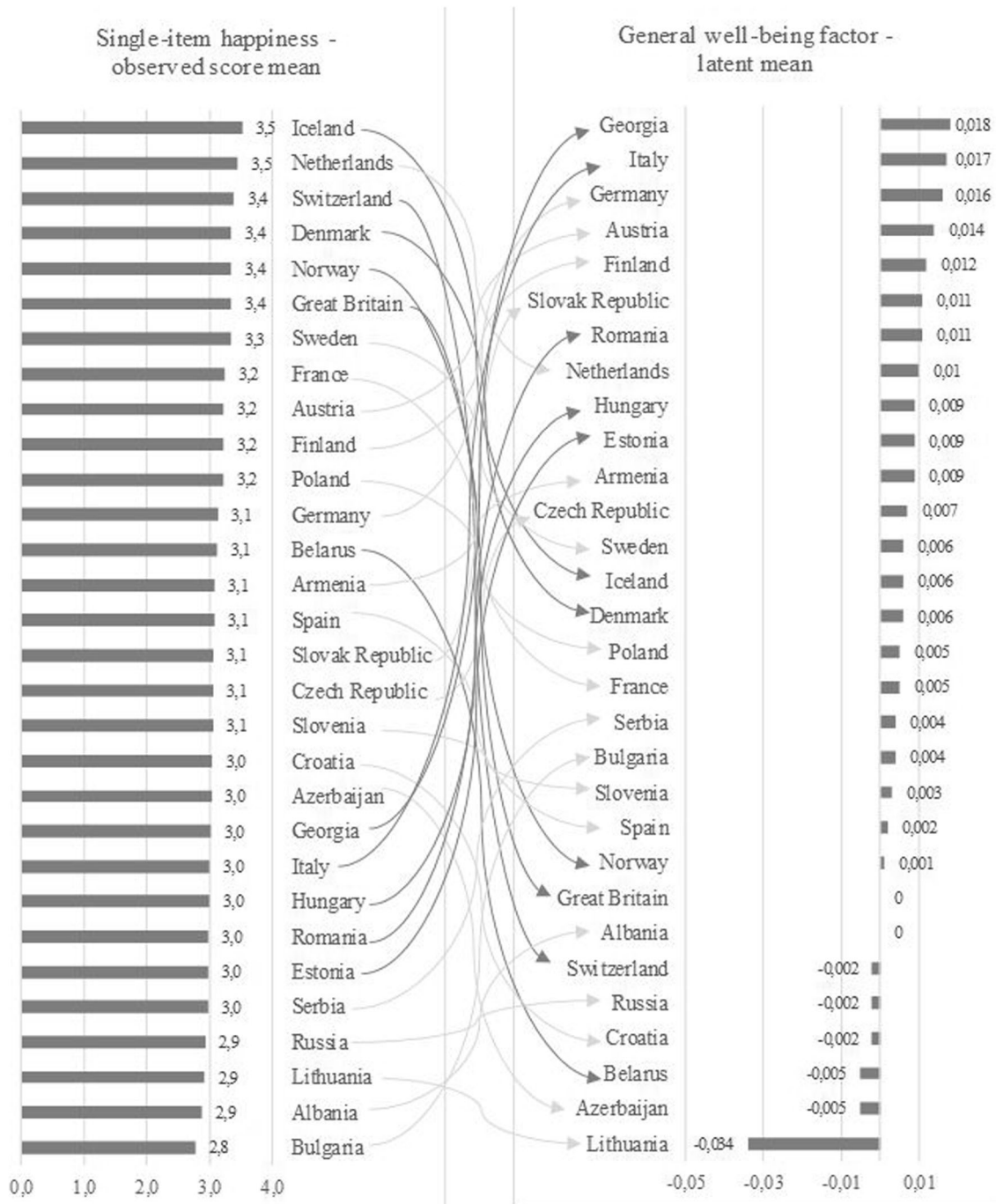
observed score mean rankings of around 0.2 (Fig. 5). The lower correlation value reflects a significant change in the ranking of many countries, which were ranked five or more places above or below based on the latent mean, as opposed to the ranking based on the observed score mean of happiness.

**Discussion and conclusion**

Due to the lack of appropriate multi-item instruments for measuring general well-being constructs in cross-national surveys, the use of single-item measures still predominates. The cultural

invariance of these measures is difficult to assess empirically because the same item has to be repeated in the survey using three different methods (Revilla and Saris, 2011). Large sample surveys do not use this multimethod design, so researchers usually assume that the reported levels of happiness and life satisfaction are comparable across countries (Jovanović and Brdar, 2018).

Thus, this study aimed to introduce an innovative approach to assess the measurement invariance of single-item measures across countries based on the post-hoc synthesis of multi-item instruments from several theoretically and empirically suitable single-item

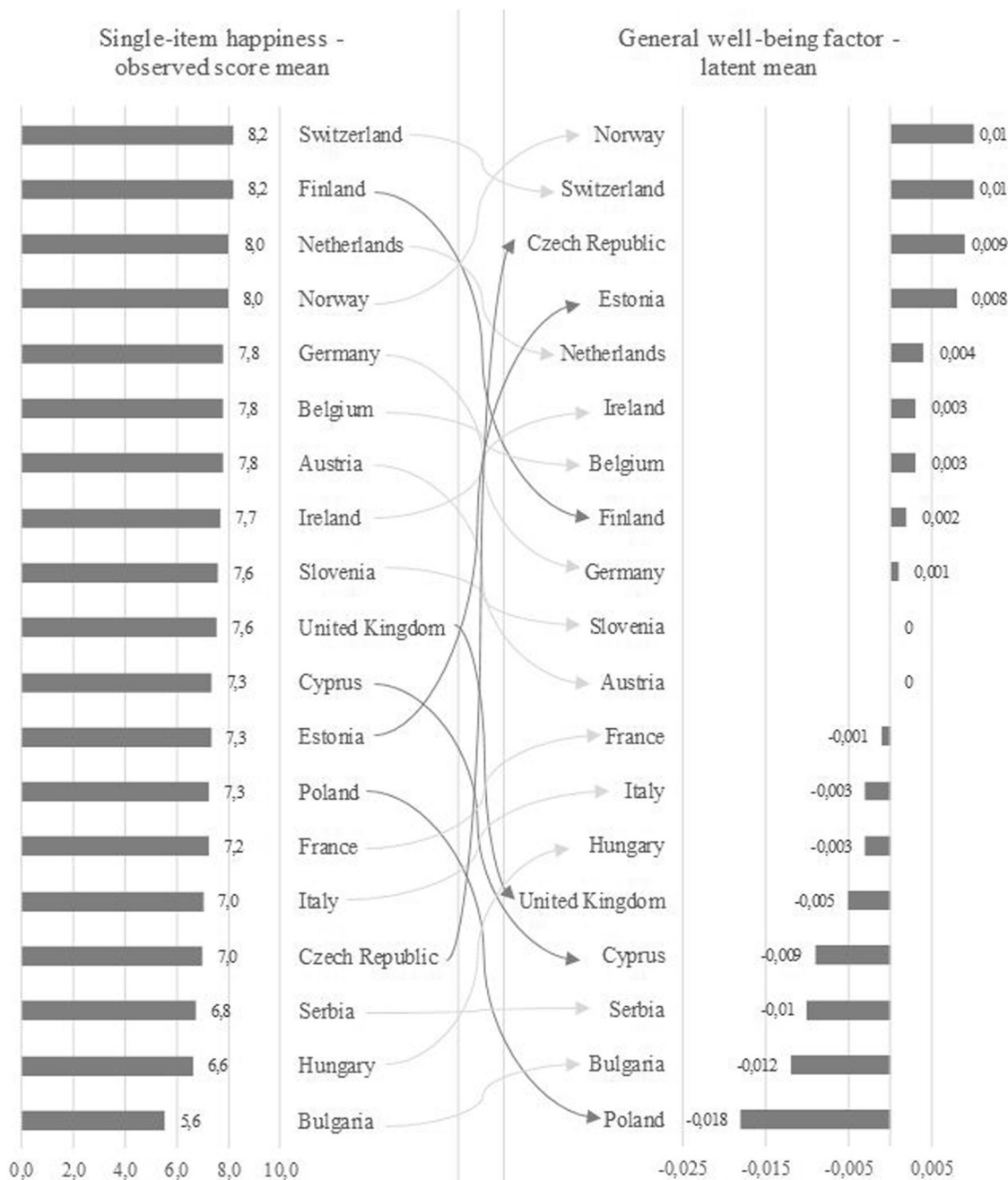


**Fig. 3 Latent and observed score means differences for general well-being factor and single-item measures of happiness/European Values Study 2017.** Data source: European Value Study 2017; author's figure. Note: Single-item question: Taking all things together, would you say you are [how happy would you say you are]? (1 not at all happy-4 very happy). Latent means were estimated in approximate invariant model (with a prior variance of 0.001). Pearson correlation coefficient between rankings of countries result from latent and observed score means comparison. single-item happiness score = 0.15, single-item life satisfaction score = 0.24. Observed scores that should not be used for comparison due to non-invariance are highlighted in black. Single grey arrows indicate the different ranking of a particular country by the observed score mean of happiness versus the ranking of country by latent mean. Light grey arrows indicate a small change in the ranking of a given country, dark grey arrows indicate a significant change in the ranking of a given country.

measures. To conduct analyses of the invariance of single-item measures of happiness and life satisfaction across six different survey programs, 45 data samples from 1976 to 2018 were used to increase the robustness and credibility of the conclusions presented. To assess the extent of specific single-item non-invariance, we used the latest Bayesian approximation approach.

Based on the results, we can conclude that the single-item happiness measure showed slightly better measurement

invariance results than the life satisfaction measure across multiple survey programs and time points. Overall, the factor loadings/intercepts of the happiness item deviated to a lesser extent and thus showed comparability across more countries. Moreover, no significant differences were found between the verbal and numerical scales of this item. The single-item happiness measure showed better results in terms of cross-cultural comparability, regardless of whether it was used with a four-point verbal scale, a

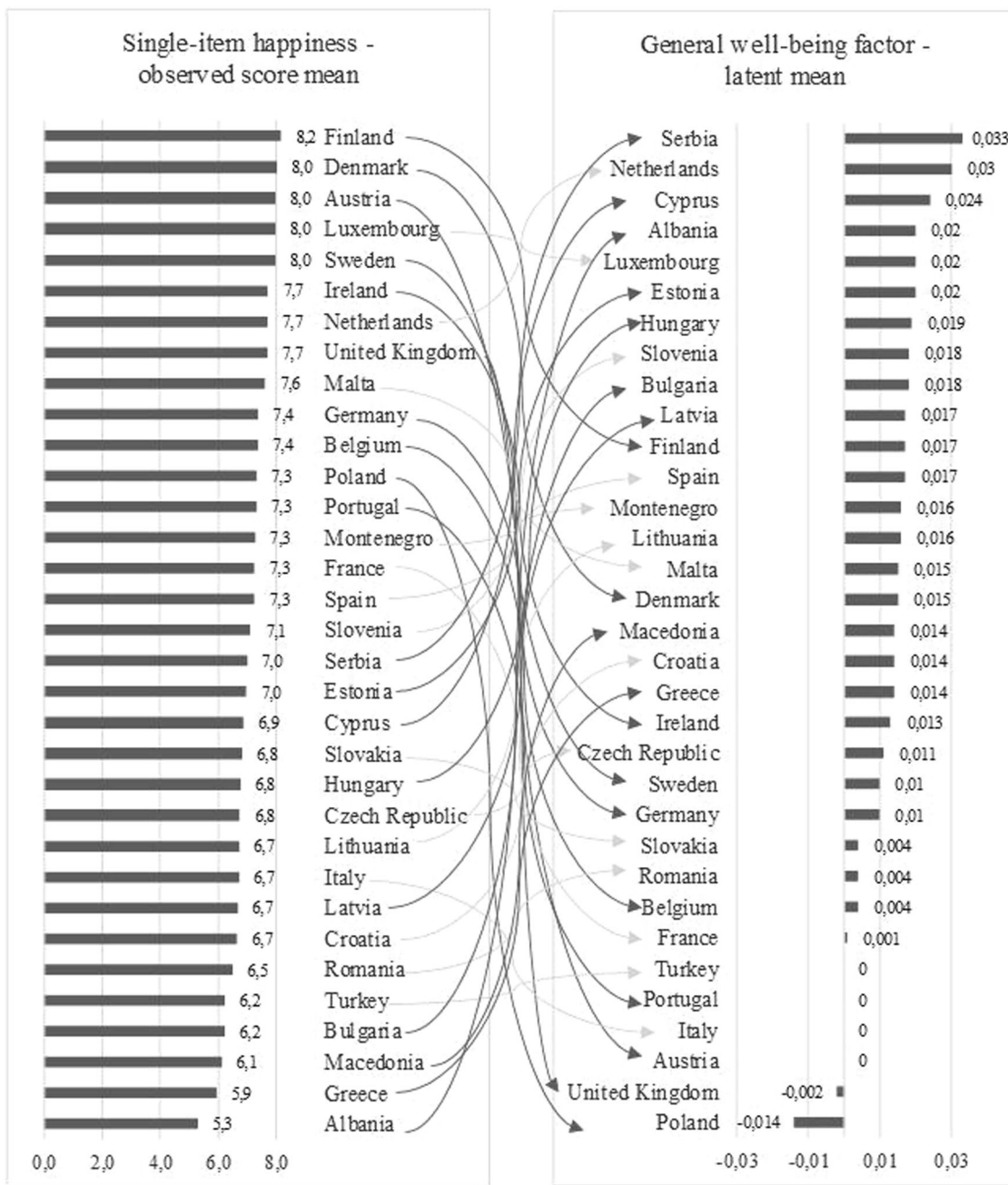


**Fig. 4 Latent and observed score means differences for general well-being factor and single-item measures of happiness/European Social Survey 2018.** Data source: European Social Survey 2018; author's figure. Note: Single-item question: Taking all things together, would you say you are [how happy would you say you are]? (0 extremely unhappy, 10 extremely happy). Latent means were estimated in approximate invariant model (with a prior variance of 0.001). Pearson correlation coefficient between rankings of countries result from latent and observed score means comparison. single-item happiness score = 0.57, single-item life satisfaction score = 0.6. Observed scores that should not be used for comparison due to non-invariance are highlighted in black. Single arrows indicate the different ranking of a particular country by the observed score mean of happiness versus the ranking of country by latent mean. Light grey arrows indicate a small change in the ranking of a given country, dark grey arrows indicate a significant change in the ranking of a given country.

seven-point verbal scale, or a 10-point numerical scale, and regardless of the modified wording in the ISSP surveys. This also indicated that the construct of happiness could be more culturally universal than the construct of life satisfaction. In our view, if researchers are forced to include only one general well-being item in the cross-national questionnaire, or if they need to analyse only a single well-being item, they might prefer to choose the happiness item rather than the life satisfaction item, which has a higher chance of being comparable across countries. Conversely, the

parameters of the subjective health or life satisfaction items were highly variable, showed less comparability across countries, and should not be included in cross-national questionnaires as single-item measures but rather as part of multi-item scales.

The assessment of specific item non-invariance showed that the single-item measures of happiness and life satisfaction were approximately metric and scalar invariant across most survey programs, rounds, and countries, which is a sufficient level of measurement invariance for analyses comparing their item



**Fig. 5 Latent and observed score means differences for general well-being factor and single-item measures of happiness/European Quality of Life Survey 2016.** Data source: European Quality of Life Survey 2016; author's figure. Note: Single-item question: Taking all things together, would you say you are [how happy would you say you are]? (1 very unhappy, 10 very happy). Latent means were estimated in approximate invariant model (with a prior variance of 0.001). Pearson correlation coefficient between rankings of countries result from latent and observed score means comparison. single-item happiness score = 0.1, single-item life satisfaction score = -0.17. Observed scores that should not be used for comparison due to non-invariance are highlighted in black. Single arrows indicate the different ranking of a particular country by the observed score mean of happiness versus the ranking of country by latent mean. Light grey arrows indicate a small change in the ranking of a given country, dark grey arrows indicate a significant change in the ranking of a given country.

covariances, unstandardized regression coefficients, and item means across these samples (Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000). However, the results revealed a problematic use of the observed item means in certain countries. The item parameters of the single-item happiness/life satisfaction measure deviated in several countries in each round of a specific cross-national survey. Thus, only approximate partial scalar invariance held for these single items; therefore, they could not be compared across all participating countries but only between a few of them (more details in the Results section).

For example, in the WVS 2017, the item parameters of the two single-item measures deviated significantly in South Korea, Germany, Egypt, and Cyprus. Neither the single-item measure of happiness nor that of life satisfaction should be compared across these countries based on the average mean of the observed scores because “it has not been confirmed that the respondents with the same value on the construct have the same expected response, irrespective of the group they belong to” (Davidov et al. 2014). The pattern of measurement invariance violations across countries was not random, as certain countries tended to have

measurement invariance violations for multiple surveys and instruments. However, the list of these countries is so diverse that no clear regional trends can be identified. For example, we cannot say that researchers can make comparisons of well-being construct between European and Latin American countries but not for some African countries. It depends on the specific survey program and round.

Therefore, in general, the observed score means can be used for cross-country comparisons, but the measurement invariance of the items must be empirically tested, not just assumed, as this is a prerequisite for meaningful cross-country analysis (Raudenská, 2020). For single-item instruments, we recommend using the proposed alternative approach of creating a synthetic (post-hoc) multi-item instrument to test measurement invariance and then excluding from the comparison the countries where the item intercepts differ significantly. For multi-item instruments, researchers may follow established practices. Whenever possible, researchers should still use verified multi-item scales rather than single-item measures for cross-national comparisons, and latent means should be preferred when it is possible to use them. The results will be more reliable.

It is beyond the scope of this study to discuss a possible explanation for the lack of comparability between countries in certain general well-being items. The survey and data documentation did not reveal any specifics related to the poor translation of questions or response scales or to changes in the data collection or research design across countries. However, the implications of different items functioning in specific cultural contexts should be explored in detail. Oishi et al. (2013) pointed out that there are cultural differences in the concept of happiness and that the meanings of happiness and satisfaction might differ across cultures, therefore affecting survey responses across cultures. Regarding the issue of different meanings and problematic translations associated with an item, Bjørnskov (2010) mentioned that the Russian, English, and French translations of the word “happy” mean both happy and lucky, while the Danish translation of the word “lykkelig” and the German word “glücklich” puts more emphasis on achievement and refers to something stronger than just being “happy” (see also Wierzbicka, 1999; Lolle and Andersen, 2016). Similarly, in Slavic languages, the word “happy” has a much more restricted meaning: “it is generally reserved for rare states of profound bliss, or total satisfaction with serious things such as love, family, the meaning of life, and so on” (Barańczak, 1990, p. 12). Köse (2015) made an interesting argument that happiness is a typical Western concept. Unfamiliarity with it in non-Western countries could lead to misunderstandings. For example, the words used in other languages to translate the English words “happy” and “satisfaction” may not exactly match, and cross-national differences may be partly artefacts of language (Wierzbicka, 2004).

Regarding the response categories of the scale, which could also cause the items to function culturally differently, Benítez et al. (2018) asserted that “in the happiness/life satisfaction scale, there are no differences in the interpretation of the extreme categories of response options; however, significant differences appear in the interpretation of intermediate categories.” Finally, Wierzbicka (2004) raised the very important question of whether it is true that nations differ in happiness or whether they differ in what they are willing to report about their state of happiness. Qualitative approaches, such as cognitive interviews or web probing, or the application of item response theory, offer a complementary tool to explain the reasons behind the different reported happiness and the non-invariance of items in a given country (see Meitinger, 2017, for more).

Our proposed approach, based on the synthesis of multi-item instruments, should be used with the utmost caution. We assessed the measurement invariance of the selected items using other

subjective well-being measures available in specific cross-national questionnaires to create the best possible general well-being (multi-item) construct. We made this decision both on a theoretical basis, because they reflect different facets of well-being similar to the well-established multi-item scales, and on an empirical basis, because a high correlation between the items confirmed that the selected items appeared to be the best choice. We encourage other researchers to use theoretical and empirical arguments to support their decision to select specific single items and to statistically examine the potential contamination of latent construct measurement using random correlations. We believe that the scope of our analysis has shed light on the degree of approximate measurement invariance of single-item measures of well-being across the world and supports caution against studies that rank countries according to their level of happiness or life satisfaction without any evidence of the level of comparability achieved.

### Data availability

The datasets analysed during the current study are available in the [JD Systems Institute & WVSA Secretariat] repository, <http://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>; <https://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>; in the [GESIS Data Archive] repository, [<https://doi.org/10.4232/1.11564>; <https://doi.org/10.4232/1.12252>; <https://doi.org/10.4232/1.12661>; <https://doi.org/10.4232/1.13322>; <https://doi.org/10.4232/1.12253>; <https://doi.org/10.4232/1.13560>; <https://doi.org/10.4232/1.10074>]; in the [Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC] repository, [<https://doi.org/10.21338/NSD-ESS-CUMULATIVE>; <https://doi.org/10.21338/NSD-ESS9-2018>].

Received: 1 December 2022; Accepted: 23 October 2023;

Published online: 23 November 2023

### Notes

- One of the valuable exception is, for example, the section of methodology testing in the European Social Survey. The ESS Core Scientific Team undertakes a range of activities related to data quality assessment across ESS rounds; these include also evaluating the quality and comparability of its measurement instruments. See <https://www.europeansocialsurvey.org/methodology/>.
- Supplementary Information are also available online at: [https://www.soc.cas.cz/sites/default/files/publikace/supplementary\\_information.pdf](https://www.soc.cas.cz/sites/default/files/publikace/supplementary_information.pdf).
- <http://www.oecd.betterlifeindex.org/#/111111111111>.
- <http://worldhappiness.report/ed/2018/>.
- <https://mobilityexchange.mercer.com/Insights/quality-of-living-rankings>.
- Veenhoven, R. (2004). World Database of Happiness: Continuous register of research on subjective appreciation of life. Retrieved from <http://hdl.handle.net/1765/8683>.
- Australia=1813, New Zealand=1057.
- ISSP 2002: Austria = 1352, Great Britain = 1960, Northern Ireland = 987, US = 1171, Netherlands = 1249, Norway = 1475, Sweden = 1080, New Zealand = 1025, France = 1903, Denmark = 1379, Belgium/Flanders = 1360, Finland = 1353/ISSP 2011: Australia = 1946, Belgium = 3083, Denmark = 1388, Finland = 1340, France = 3319, Germany (part) = 1409, Italy = 1186, Japan = 1306, Netherlands = 1472, Norway = 1834, Sweden = 1158, United Kingdom = 936, US (partly) = 311/ISSP 2012: Australia = 1612, Canada = 983, Denmark = 1403, Finland = 1171, France = 2409, Germany (partly) = 1464, Iceland = 246, Ireland = 1215, Japan = 1212, Norway = 1444, Sweden = 1060, United Kingdom = 950, US (partly) = 164/ISSP 2017: Australia = 1317, Denmark = 1079, Finland = 1074, France = 1489, Germany (partly) = 1455, Iceland = 1450, Japan = 1609, New Zealand = 1357, Surinam (partly) = 536, Sweden = 1125, United Kingdom = 1595, US (partly) = 111.
- Denmark (partly) = 1666, Finland (partly) = 811, Germany (partly) = 3237, Iceland (partly) = 1591, Netherlands (partly) = 2035, Switzerland (partly) = 2987.

### References

- Abdel-Khalek AM (2006) Measuring happiness with a single-item scale. *Soc Behav Personal* 34(2):139–150. <https://doi.org/10.2224/sbp.2006.34.2.139>

- Atroszko P, Sawicki A, Małkonia A, Atroszko B (2017) Further validation of single-item self-report measure of satisfaction with life. In: 7th Biannual CER Comparative European Research Conference: International Scientific Conference for Ph. D. students of EU countries. Science Publishing
- Asparouhov T, Muthén B (2010) Bayesian analysis of latent variable models using Mplus. Muthén & Muthén, Los Angeles, CA, [www.statmodel.com](http://www.statmodel.com). Technical report
- Barańczak S (1990) *Breathing Under Water and Other East European Essays*. Harvard University Press, Cambridge, Mass
- Benítez I, Padilla JL, van de Vijver F, Cuevas A (2018) What cognitive interviews tell us about bias in cross-cultural research: An illustration using quality-of-life items. *Field Methods* 30(4):277–294. <https://doi.org/10.1177/1525822x18783961>
- Billiet J (2003) Cross-cultural equivalence with structural equation modeling. In: Harkness JA, Van de Vijver FJR, Mohler PPH (Eds.) *Cross-cultural survey methods*. John Wiley & Sons, pp. 247–263
- Bjørnskov C (2010) How comparable are the Gallup World Poll life satisfaction data? *J Happiness Stud* 11(1):41–60. <https://doi.org/10.1007/s10902-008-9121-6>
- Bujacz A, Vittersø J, Huta V, Kaczmarek LD (2014) Measuring hedonia and eudaimonia as motives for activities: Cross-national investigation through traditional and Bayesian structural equation modeling. *Front Psychol* 5:984. <https://doi.org/10.3389/fpsyg.2014.00984>
- Cernat A, Revilla M (2021) Moving from face-to-face to a web panel: Impacts on measurement quality. *J Surv Stat Methodol* 9:745–763
- Ciecuch J, Davidov E, Algesheimer R, Schmidt P (2018) Testing for approximate measurement invariance of human values in the European Social Survey. *Sociol Methods Res* 47(4):665–686. <https://doi.org/10.1177/0049124117701478>
- Ciecuch J, Davidov E, Schmidt P, Algesheimer R, Schwartz SH (2014) Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Front Psychol* 5:1–10. <https://doi.org/10.3389/fpsyg.2014.00982>
- Checa I, Perales J, Espejo B (2019) Measurement invariance of the Satisfaction with Life Scale by gender, age, marital status, and educational level. *Qual Life Res* 28(4):963–968. <https://doi.org/10.1007/s11136-018-2066-2>
- Cheung F, Lucas RE (2014) Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Qual Life Res* 23(10):2809–2818. <https://doi.org/10.1007/s11136-014-0726-4>
- Clench-Aas J, Nes RB, Dalgard OD, Aarø LE (2011) Dimensionality and measurement invariance in the Satisfaction With Life Scale in Norway. *Qual Life Res* 20(8):1307–1317. <https://doi.org/10.1007/s11136-011-9859-x>
- Cummins RA, Eckersley R, Pallant J, Van Vugt J, Misajon R (2003) Developing a national index of subjective well-being: The Australian Unity Well-being Index. *Soc Indic Res* 64(2):159–190. <https://doi.org/10.1023/a:1024704320683>
- Davidov E, Ciecuch J, Meuleman B, Schmidt P, Algesheimer R, Hausherr M (2015) The comparability of measurements of attitudes toward immigration in the European Social Survey. *Pub Opin Q* 79(S1):244–266. <https://doi.org/10.1093/poq/nfv008>
- Davidov E, Ciecuch J, Schmidt P (2018) The cross-country measurement comparability in the immigration module of the European Social Survey 2014–15. *Surv Res Methods* 12(1):15–27. <https://doi.org/10.18148/srm/2018.v12i1.7212>
- Davidov E, Meuleman B, Ciecuch J, Schmidt P, Billiet J (2014) Measurement equivalence in cross-national research. *Annu Rev Sociol* 40(1):55–75. <https://doi.org/10.1146/annurev-soc-071913-043137>
- De Beuckelaer A, Lievens F (2009) Measurement equivalence of paper-and-pencil and Internet organisational surveys: A large scale examination in 16 countries. *Appl Psychol* 58(2):336–361
- Diener ED (2009) *The science of well-being. The collected works of Ed Diener*. Springer
- Diener EDD, Chan MY (2011) Happy people live longer: Subjective well-being contributes to health and longevity. *Appl Psychol* 3(1):1–43. <https://doi.org/10.1111/j.1758-0854.2010.01045.x>
- Diener ED, Emmons RA, Larsen RJ, Griffin S (1985) The satisfaction with life scale. *J Personal Assess* 49(1):71–75. [https://doi.org/10.1207/s15327752jpa4901\\_13](https://doi.org/10.1207/s15327752jpa4901_13)
- Diener ED, Inglehart R, Tay L (2013) Theory and validity of life satisfaction scales. *Soc Indic Res* 112(3):497–527. <https://doi.org/10.1007/s11205-012-0076-y>
- Diener ED, Lucas RE, Oishi S (2018) Advances and open questions in the science of subjective well-being. *Collabra* 4(1):15. <https://doi.org/10.1525/collabra.115>
- Dimitrova R, Buzea C, Abubakar A, Stefanel D (2016) Measurement invariance of the satisfaction with life scale in Roma minority and Romanian majority adolescents in Romania. *Psihologia Resurselor Umane* 14(2):111–121. <https://doi.org/10.24837/pru.2016.2.2>
- Emerson SD, Guhn M, Gadermann AM (2017) Measurement invariance of the Satisfaction with Life Scale: Reviewing three decades of research. *Qual Life Res* 26(9):2251–2264. <https://doi.org/10.1007/s11136-017-1552-2>
- Esnola I, Benito M, Agirre IA, Freeman J, Sarasa M (2017) Measurement invariance of the Satisfaction with Life Scale (SWLS) by country, gender and age. *Psicothema* 29(4):596–601. <https://doi.org/10.1007/s11136-019-02224-7>
- European Social Survey Cumulative File, ESS 1-8 (2018). Data file edition 1.0. NSD - Norwegian Centre for Research Data, Norway - Data Archive and distributor of ESS data for ESS ERIC. <https://doi.org/10.21338/NSD-ESS-CUMULATIVE>
- ESS Round 9: European Social Survey Round 9 Data (2018). Data file edition 2.0. NSD - Norwegian Centre for Research Data, Norway - Data Archive and distributor of ESS data for ESS ERIC. <https://doi.org/10.21338/NSD-ESS-2018>
- EVS (2015): European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008). GESIS Data Archive, Cologne. ZA4804 Data file Version 3.0.0, <https://doi.org/10.4232/1.12253>
- EVS (2020): European Values Study 2017: Integrated Dataset (EVS 2017). GESIS Data Archive, Cologne. ZA7500 Data file Version 4.0.0, <https://doi.org/10.4232/1.13560>
- Fonberg J, Smith AP (2019) The validity of a single question about life satisfaction. *Int J Arts Humanit Soc Sci* 4:38–44
- Fors F, Kulin J (2016) Bringing affect back in: Measuring and comparing subjective well-being across countries. *Soc Indic Res* 127(1):323–339. <https://doi.org/10.1007/s11205-015-0947-0>
- Haerpfer C, Inglehart R, Moreno A, Welzel C, Kizilova K, Diez-Medrano J, Lagos M, Norris P, Ponarin E, Puranen B, et al., (Eds.) (2020) *World Values Survey: Round Seven—Country-pooled Datafile*. JD Systems Institute & WVSA Secretariat, Version: 4.0.0, Madrid, Spain & Vienna, Austria, <http://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>
- Horn JL, McArdle JJ (1992) A Practical and Theoretical Guide to Measurement Invariance in Aging Research. *Exp Aging Res* 18(3):117–144. <https://doi.org/10.1080/03610739208253916>
- Hox JJ, de Leeuw ED, Zijlman EA (2015) Measurement equivalence in mixed mode surveys. *Front Psychol* 6:87
- Huppert FA, Marks N, Clark A, Siegrist J, Stutzer A, Vittersø J, Wahrendorf M (2009) Measuring well-being across Europe: Description of the ESS well-being module and preliminary findings. *Soc Indic Res* 91(3):301–315. <https://doi.org/10.1007/s11205-008-9346-0>
- Huppert FA, So TTC (2013) Flourishing across Europe: Application of a New Conceptual Framework for Defining Well-Being. *Soc Indic Res* 110(3):837–861. <https://doi.org/10.1007/s11205-011-9966-7>
- Inglehart R, Haerpfer C, Moreno A, Welzel C, Kizilova K, Diez-Medrano J, Lagos M, Norris P, Ponarin E, Puranen B, et al., (Eds.) (2014) *World Values Survey: Round Six—Country-Pooled Datafile*. JD Systems Institute, Version: v2018-09-12, Madrid, Spain & Vienna, Madrid, <https://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>
- ISSP Research Group (2013) *International Social Survey Programme: Family and Changing Gender Roles III—ISSP 2002*. GESIS Data Archive, ZA3880 Data file Version 1.1.0, Cologne, 10.4232/1.11564
- ISSP Research Group (2015) *International Social Survey Programme: Health and Health Care - ISSP 2011*. GESIS Data Archive, ZA5800 Data file Version 3.0.0, Cologne, 10.4232/1.12252
- ISSP Research Group (2016) *International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012*. GESIS Data Archive, ZA5900 Data file Version 4.0.0, Cologne, 10.4232/1.12661
- ISSP Research Group (2019) *International Social Survey Programme: Social Networks and Social Resources - ISSP 2017*. GESIS Data Archive, ZA6980 Data file Version 2.0.0, Cologne, 10.4232/1.13322
- Jang S, Kim ES, Cao C, Allen TD, Cooper CL, Lapierre LM, Abarca N (2017) Measurement invariance of the satisfaction with life scale across 26 countries. *J Cross-Cult Psychol* 48(4):560–576. <https://doi.org/10.1177/0022022117697844>
- Jovanović V (2016) The validity of the Satisfaction with Life Scale in adolescents and a comparison with single-item life satisfaction measures: a preliminary study. *Qual Life Res* 25(12):3173–3180. <https://doi.org/10.1007/s11136-016-1331-5>
- Jovanović V, Brdar I (2018) The cross-national measurement invariance of the Satisfaction with Life Scale in a sample of undergraduate students. *Personal Individ Differ* 128:7–9. <https://doi.org/10.1016/j.paid.2018.02.010>
- Jovanović V, Lazić M (2020) Is longer always better? A comparison of the validity of single-item versus multiple-item measures of life satisfaction. *Appl Res Qual Life* 15(3):675–692. <https://doi.org/10.1007/s11482-018-9680-6>
- Jovanović V, Cummins RA, Weinberg M, Kaliterna L, Prizmic-Larsen Z (2018) *Personal Well-being Index: A cross-cultural measurement invariance study across four countries*. *J Happiness Stud* 20(3):1–17. <https://doi.org/10.1007/s10902-018-9966-2>
- Kahneman D, Diener ED, Schwarz N (Eds.) (1999) *Well-being: The foundations of hedonic psychology*. Russell Sage Found
- Kim ES, Cao C, Wang Y, Nguyen DT (2017) Measurement invariance testing with many groups: A comparison of five approaches. *Struct Equ Model* 24(4):524–544. <https://doi.org/10.1080/10705511.2017.1304822>
- Köse İA (2015) Psychometric properties of the Orientations to Happiness Scale and measurement invariance between samples of Turkish and Russian University



- students. *Soc Indic Res* 122(3):945–959. <https://doi.org/10.1007/s11205-014-0721-8>
- Leitgöb H, Seddig D, Asparouhov T, Behr D, Davidov E, De Roover K, ... van de Schoot R (2022) Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Soc Sci Res* 102805. <https://doi.org/10.1016/j.ssresearch.2022.102805>
- Lek K, Oberski D, Davidov E, Cieciuch J, Seddig D, Schmidt P (2018) Approximate measurement invariance. In: Johnson TP, Pennell BE, Stoop IA, Dorer B (Eds.). *Advances in comparative survey methods: Multinational, multi-regional, and multicultural contexts (3MC)*. John Wiley & Sons, pp 911–928
- Lolle HL, Andersen JG (2016) Measuring happiness and overall life satisfaction: A Danish survey experiment on the impact of language and translation problems. *J Happiness Stud* 17(4):1337–1350. <https://doi.org/10.1007/s10902-015-9646-4>
- The Mannheim Eurobarometer Trend File 1970–2002, [Computer file]. GESIS Data Archive, Cologne. GESIS Study ZA3521, 2nd. edition (2.01), Cologne, Germany, <https://doi.org/10.4232/1.10074>
- Meitinger K (2017) Necessary but insufficient. Why measurement invariance tests need online probing as a complementary tool. *Pub Opin Q* 81(2):447–472
- Meredith W (1993) Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika* 58(4):525–543. <https://doi.org/10.1007/BF02294825>
- Muthén B, Asparouhov T (2013) BSEM measurement invariance analysis. *Mplus Web Notes* 17:1–48
- Muthén L, Muthén BO (1998–2017) *Mplus user's guide* (8.). Muthén & Muthén
- Nes RB, Roysamb E, Tambs K, Harris JR, Reichborn-Kjennerud T (2006) Subjective well-being: Genetic and environmental contributions to stability and change. *Psychol Med* 36(07):1033–1042. <https://doi.org/10.1017/S0033291706007409>
- Oishi S (2012) The psychological wealth of nations: Do happy people make a happy society? Wiley-Blackwell
- Oishi S, Graham J, Kesebir S, Galinha IC (2013) Concepts of happiness across time and cultures. *Personal Soc Psychol Bull* 39(5):559–577. <https://doi.org/10.1177/0146167213480042>
- Pavot W, Diener ED, Colvin CR, Sandvik E (1991) Further validation of the Satisfaction with Life Scale: Evidence for the cross-method convergence of well-being measures. *J Personal Assess* 57(1):149–161. [https://doi.org/10.1207/s15327752jpa5701\\_17](https://doi.org/10.1207/s15327752jpa5701_17)
- Pokropek A, Schmidt P, Davidov E (2020) Choosing priors in Bayesian measurement invariance modeling: A Monte Carlo simulation study. *Struct Eq Model* 27(5):750–764. <https://doi.org/10.1080/10705511.2019.1703708>
- Ponizovsky Y, Dimitrova R, Schachner MK, van de Schoot R (2013) The satisfaction with life scale: Measurement invariance across immigrant groups. *Eur J Dev Psychol* 10(4):526–532. <https://doi.org/10.1080/17405629.2012.707778>
- Raudenská P (2020) The cross-country and cross-time measurement invariance of positive and negative affect scales: Evidence from European Social Survey. *Soc Sci Res* 86:102369. <https://doi.org/10.1016/j.ssresearch.2019.102369>
- Revilla M, Saris W (2011) How to evaluate the cross-cultural equivalence of single items. (RECSM, UPF) Mini conference: Measurement invariance: Methods, problems, and further directions. Zurich, July15–16, 2011
- Sakshaug JW, Cernat A, Silverwood RJ, Calderwood L, Ploubidis GB (2022) Measurement Equivalence in Sequential Mixed-Mode Surveys. *Surv Res Methods* 16(1):29–43
- Saris WE, Meurs A (1990) Evaluation of Measurement Instruments by Meta-Analysis of Multi-Trait Multi-Method Studies. North-Holland, Amsterdam
- Saris WE, Gallhofer I (2007) Can questions travel successfully? In: Jowell R et al. (Eds.) *Measuring attitudes cross-nationally*. Sage Publications, p 53–79. <https://doi.org/10.4135/9781849209458.n3>
- Schnettler B, Miranda-Zapata E, Lobos G, del Carmen Lapo M, Adasme-Berrios C, Hueche C (2017) Measurement invariance in the Satisfaction with Life Scale in Chilean and Ecuadorian older adults. *Personal Individ Differ* 110:96–101. <https://doi.org/10.1016/j.paid.2017.01.036>
- Seddig D, Leitgöb H (2018) Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data. *Surv Res Methods* 12(1):29–41
- Steenkamp JBEM, Baumgartner H (1998) Assessing measurement invariance in cross-national consumer research. *J Consum Res* 25(1):78–90. <https://doi.org/10.1086/209528>
- Steinmetz H (2013) Analyzing observed composite differences across groups. *Methodology* 9(1):1–12. <https://doi.org/10.1027/1614-2241/a000049>
- Strobel M, Tumasjan A, Spörrle M (2011) Be yourself, believe in yourself, and be happy: Self-efficacy as a mediator between personality factors and subjective well-being. *Scand J Psychol* 52(1):43–48. <https://doi.org/10.1111/j.1467-9450.2010.00826.x>
- Tomás JM, Gutiérrez M, Sancho P, Romero I (2015) Measurement invariance of the Satisfaction With Life Scale (SWLS) by gender and age in Angola. *Personal Individ Differ* 85:182–186. <https://doi.org/10.1016/j.paid.2015.05.008>
- Tucker KL, Ozer DJ, Lyubomirsky S, Boehm JK (2006) Testing for measurement invariance in the Satisfaction With Life Scale: A comparison of Russians and North Americans. *Soc Indic Res* 78(2):341–360. <https://doi.org/10.1007/s11205-005-1037-5>
- Van De Schoot R, Kluytmans A, Tummers L, Lugtig P, Hox J, Muthén B (2013) Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Front Psychol* 4:770. <https://doi.org/10.3389/fpsyg.2013.00770>
- Van de Velde S, Bracke P, Levecque K, Meuleman B (2010) Gender differences in depression in 25 European countries after eliminating measurement bias in the CES-D 8. *Soc Sci Res* 39(3):396–404. <https://doi.org/10.1016/j.ssresearch.2010.01.002>
- Vandenberg RJ, Lance CE (2000) A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 3(1):4–70. <https://doi.org/10.1177/109442810031002>
- Vanhoutte B, Nazroo J (2014) Cognitive, affective and eudemonic well-being in later life: Measurement equivalence over gender and life stage. *Sociol Res Online* 19(2):1–14. <https://doi.org/10.5153/sro.3241>
- Veenhoven R (2010) How universal is happiness? In: Diener E, Helliwell J, Kahneman D (Eds.). *International differences in well-being*. Oxford University Press, New York, p 328–350
- Veenhoven R (2012) Happiness: Also known as “life-satisfaction” and “subjective well-being.” In: Land KC, Michalos AC, Sirgy MJ (Eds.) *Handbook of social indicators and quality of life research*. Springer Publishers, p 63–77. [https://doi.org/10.1007/978-94-007-2421-1\\_3](https://doi.org/10.1007/978-94-007-2421-1_3)
- Waterman AS (1993) Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *J Personal Soc Psychol* 64(4):678–691. <https://doi.org/10.1037/0022-3514.64.4.678>
- Whisman MA, Judd CM (2016) A cross-national analysis of measurement invariance of the Satisfaction With Life Scale. *Psychol Assess* 28(2):239–244. <https://doi.org/10.1037/pas0000181>
- Wierzbicka A 1999 *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511521256>
- Wierzbicka A (2004) ‘Happiness’ in cross-linguistic & cross-cultural perspective. *Daedalus* 133(2):34–43. <https://doi.org/10.1162/001152604323049370>
- Zercher F, Schmidt P, Cieciuch J, Davidov E (2015) The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. approximate measurement invariance. *Front Psychol* 6:1–11. <https://doi.org/10.3389/fpsyg.2015.00733>
- Żemojtel-Piotrowska M, Piotrowski JP, Cieciuch J, Adams BG, Osin EN, Ardi R, Esteves C et al. (2017) Measurement invariance of Personal Well-being Index (PWI-8) across 26 countries. *J Happiness Stud* 18(6):1697–1711. <https://doi.org/10.1007/s10902-016-9795-0>

## Acknowledgements

This work was supported by the Czech Science Foundation “Finding health and happiness in the post-covid-19 world” (Grant No. 22-09220S). An earlier version was presented at the 10th ESRA conference at the University of Milan Bicocca (July, 2023). The author is grateful to the participants of the seminar, especially Daniel Seddig, Eldad Davidov, and Peter Schmidt.

## Author contributions

The corresponding author was responsible for all aspects of the manuscript.

## Competing interests

The author declares no competing interests.

## Ethical approval

Ethical approval was not required as the study did not involve any studies with human participants performed by any of the authors.

## Informed consent

Informed consent was not required as the study did not involve any studies with human participants performed by any of the authors.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-02299-1>.

**Correspondence** and requests for materials should be addressed to Petra Raudenská.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© Institute of Sociology, Czech Academy of Sciences 2023